

The Draft Genome of an Octocoral, *Dendronephthya gigantea*

Yeonsu Jeon^{1,2,†}, Seung Gu Park^{1,†}, Nayun Lee^{3,†}, Jessica A. Weber^{4,5}, Hui-Su Kim¹, Sung-Jin Hwang⁶, Seonock Woo⁷, Hak-Min Kim^{1,2}, Youngjune Bhak^{1,2}, Sungwon Jeon^{1,2}, Nayoung Lee³, Yejin Jo³, Asta Blazyte¹, Taewoo Ryu⁸, Yun Sung Cho^{1,2,9}, Hyunho Kim¹⁰, Jung-Hyun Lee⁷, Hyung-Soon Yim⁷, Jong Bhak^{1,2,9,10,*,‡}, and Seungshic Yum^{3,11,*,‡}

¹Korean Genomics Industrialization and Commercialization Center (KOGIC), Ulsan National Institute of Science and Technology (UNIST), Ulsan, Republic of Korea

²Department of Biomedical Engineering, School of Life Sciences, Ulsan National Institute of Science and Technology (UNIST), Ulsan, Republic of Korea

³Ecological Risk Research Division, Korea Institute of Ocean Science and Technology (KIOST), Geoje, Republic of Korea

⁴Department of Genetics, Harvard Medical School, Boston, Massachusetts

⁵Department of Biology, University of New Mexico

⁶Department of Life Science, Woosuk University, Republic of Korea

⁷Marine Biotechnology Research Center, Korea Institute of Ocean Science and Technology (KIOST), Busan, Republic of Korea

⁸APEC Climate Center, Busan, South Korea

⁹Clinomics Inc., Ulsan, Republic of Korea

¹⁰Personal Genomics Institute, Genome Research Foundation, Cheongju, Republic of Korea

¹¹Faculty of Marine Environmental Science, University of Science and Technology (UST), Geoje, Republic of Korea

[†]These authors contributed equally to this work.

[‡]These authors jointly supervised this work.

*Corresponding authors: E-mails: jongbhak@genomics.org; syum@kiost.ac.kr.

Accepted: February 26, 2019

Data deposition: The octocoral whole genome and transcriptome project has been deposited at DDBJ/ENA/GenBank under the accession PRJNA507923 and PRJNA507943. DNA and RNA sequencing reads have been uploaded to the NCBI Read Archive under the accession (SRR8293699, and SRR8293698 and SRR8293935, and SRR8293936), respectively. The genome assembly has been deposited at DDBJ/ENA/GenBank under the accession RSEI01000000.

Abstract

Coral reefs composed of stony corals are threatened by global marine environmental changes. However, soft coral communities of octocorallian species, appear more resilient. The genomes of several cnidarians species have been published, including from stony corals, sea anemones, and hydra. To fill the phylogenetic gap for octocoral species of cnidarians, we sequenced the octocoral, *Dendronephthya gigantea*, a nonsymbiotic soft coral, commonly known as the carnation coral. The *D. gigantea* genome size is ~276 Mb. A high-quality genome assembly was constructed from PacBio long reads (29.85 Gb with 108× coverage) and Illumina short paired-end reads (35.54 Gb with 128× coverage) resulting in the highest N50 value (1.4 Mb) reported thus far among cnidarian genomes. About 12% of the genome is repetitive elements and contained 28,879 predicted protein-coding genes. This gene set is composed of 94% complete BUSCO ortholog benchmark genes, which is the second highest value among the cnidarians, indicating high quality. Based on molecular phylogenetic analysis, octocoral and hexacoral divergence times were estimated at 544 MYA. There is a clear difference in *Hox* gene composition between these species: unlike hexacorals, the Antp superclass *Evx* gene was absent in *D. gigantea*. Here, we present the first genome assembly of a nonsymbiotic octocoral, *D. gigantea* to aid in the comparative genomic analysis of cnidarians, including stony and soft corals, both symbiotic and nonsymbiotic. The *D. gigantea* genome may also provide clues to mechanisms of differential coping between the soft and stony corals in response to scenarios of global warming.

Key words: soft coral, genome, octocoral, nonsymbiotic coral, cnidarian, *Dendronephthya gigantea*.

Introduction

Corals that belong to the phylum Cnidaria, class Anthozoa, provide habitats for a diversity of marine organisms (Friedlander and Parrish 1998) and are foundational members of the benthic community playing a major role in energy transfer between plankton and the benthos (van de Water et al. 2018). Corals capture large quantities of plankton and thereby regulate the primary and secondary production of the coastal food chains (Gili and Coma 1998; van de Water et al. 2018). Corals can be classified into hexacorals (stony corals and sea anemones) and octocorals (soft corals and sea fans). Global marine environmental changes, represented by seawater temperature rise and ocean acidification, are known to threaten coral reefs consisting of stony corals in tropical regions (Hoegh-Guldberg and research 1999; Carpenter et al. 2008). However, soft coral communities in temperate and subtropical regions, seem to prosper owing to their ability to disperse north as distribution limits extend (de Paula and Creed 2004; Santodomingo et al. 2013). To date much research has focused on understanding stony coral susceptibility to coral bleaching (Hoegh-Guldberg and research 1999) due to global warming and ocean acidification (Ries et al. 2010; de Putron et al. 2011; Pandolfi et al. 2011; Inoue et al. 2013). Yet, soft corals, which have sclerites, are less vulnerable to such environmental changes (Inoue et al. 2013) and it is suggested that temperate shallow-living octocorals are able to withstand increased levels of temperature and acidification (Lopes et al. 2018). Though, given the significant biological differences between the stony and soft corals in terms of calcification and survival strategies in the changing environment, only hexacoral genomes have been sequenced and analyzed (Putnam et al. 2007; Shinzato et al. 2011; Baumgarten et al. 2015; Snelling et al. 2017; Voolstra et al. 2017; Ying et al. 2018). Moreover, it is also beneficial to add an octocoral special to help understand the already available hexacoral genomes.

Here, we report the first genome assembly of an octocoral, *Dendronephthya gigantea*, commonly known as carnation coral. *D. gigantea* is a dominant species in the most southern coastal part of Korea (Hwang and Song 2007), in temperate and subtropical regions where yearly water temperature ranges from 14 °C to 26 °C (Hwang and Song 2007). In general, colonies of this species inhabit shallow water from 10 to 20 m in depth. It is an independent nonsymbiotic gonochoric internal brooder. It preys on zooplankton and phytoplankton and does not possess zooxanthellae (Imbs et al. 2007) in contrast to reef-building *Acropora* species. Our draft genome may therefore serve as a resource for evolutionary studies of azooxanthellate octocorals in terms of understanding different coping strategies mediating against rapid environmental changes in comparison to published stony coral genomes.

Sequencing and De Novo Genome Assembly

We estimated the genome size of *D. gigantea* to be 276 Mb (276,273,039 bp) using Illumina HiSeq 2500 short paired-end reads (35.54 Gb with 128-fold coverage) of at a k-mer size of 17. The graph for the k-mer frequency distribution showed that there were two peaks and the heterozygosity of the *D. gigantea* genome is high (Liu et al. 2013) (supplementary fig. 1, Supplementary Material online). This finding is consistent with previous reports of invertebrates showing relatively high levels of genome heterozygosity (Ellegren and Galtier 2016).

We used PacBio long reads (29.85 Gb with 108-fold coverage) for an initial draft assembly which is complemented by Illumina short paired-end reads (35.54 Gb with 128-fold coverage) for error-correction. Bacterial and fungal DNA reads (1.18%) were filtered out during genome assembly. The final assembly resulted in a 286 Mb genome, which covers 103.58% of the estimated genome size of 276 Mb (Table 1). The final contig N50 value achieved was 1,445,523 bp (Table 1). The *D. gigantea* genome assembly produced has the longest N50 length (1.4 Mb) reported among cnidarian genomes thus far (Table 1). In addition, the self-mapping rate of Illumina short paired-end reads to the genome assembly was very high (95.9%).

Gene Prediction, Annotation, and Quality Assessment

We found close to 29,000 protein-coding genes in *D. gigantea* using two different methodologies (see supplementary table 1, Supplementary Material online). The first and second approach predicted 28,879 and 28,937 protein-coding genes in the *D. gigantea* genome, respectively.

We compared both gene sets using BUSCO (version 3.0.2) (Simão et al. 2015; Waterhouse et al. 2018) which showed comparable high quality, increasing our confidence in the predicted gene set. The gene set obtained by the first method showed a slightly higher quality (93.97% complete BUSCO genes) than that of the second method (93.35%) (supplementary table 2, Supplementary Material online).

The *D. gigantea* gene set was of high quality among the cnidarians and covered ~94% of the complete BUSCO ortholog benchmark genes (supplementary fig. 2, Supplementary Material online). We compared the quality of the *D. gigantea* gene models with six published cnidarians (*Aiptasia pallida*, *Acropora digitifera*, *Hydra magnipapillata*, *Nematostella vectensis*, *Orbicella faveolata*, and *Stylophora pistillata*). The *D. gigantea* gene models had ~87% complete single copy BUSCO genes (supplementary fig. 2, Supplementary Material online). It also had the second highest value of complete BUSCO genes (~94%) which included both single copy and duplicated genes among cnidarians (supplementary fig. 2, Supplementary Material online).

Almost 12% of the *D. gigantea* genome consists of repeat elements. We found transposable elements make up an 11.97% of the *D. gigantea* genome, in which tandem repeats

Table 1Statistics of the *Dendronephthya gigantea* Genome Assembly Compared to Other Cnidarians

	<i>Dendronephthya gigantea</i>	<i>Orbicella faveolata</i>	<i>Stylophora pistillata</i>	<i>Acropora digitifera</i>	<i>Aiptasia pallida</i>	<i>Nematostella vectensis</i>	<i>Hydra magnipapillata</i>
No. of sequences	1,323	1,933	5,688	2,421	4,312	10,804	20,916
Total bases (bp)	286,131,912	485,548,939	400,120,318	447,497,157	256,132,296	356,613,585	852,170,992
Average length (bp)	216,275	251,189	70,345	184,839	59,400	33,008	40,743
SD (bp)	596,503	541,789	193,436	280,650	169,768	149,438	58,784
N50 (bp)	1,445,523	1,162,446	457,453	483,559	442,145	472,588	96,317
GC contents	37%	39%	39%	39%	36%	41%	28%

and long terminal repeat elements (LTR) represented 7.24% and 2.25% of the genome, respectively (supplementary table 3, Supplementary Material online).

Phylogenetic Analysis and *Hox* Gene Clusters Identification

We found that *D. gigantea* has diverged the earliest among the anthozoans based on our calculations. We identified that *D. gigantea* contains 12,597 orthologous gene families, excluding singletons, and 3,656 of them are shared with stony corals (*Orbicella faveolata*, *Stylophora pistillata*, and *Acropora digitifera*) and hydra (*Hydra magnipapillata*) (supplementary fig. 3, Supplementary Material online). A total of 4,863 gene families were *D. gigantea*-specific (supplementary fig. 3, Supplementary Material online). Second, we use molecular phylogenetic analysis to show that the octocoral, *D. gigantea*, is positioned between hexacorallia and hydrozoa (fig. 1A), implying that the octocoral is the earliest diverged group among anthozoans. Divergence time estimation analysis suggested the divergence of the octocoral (*D. gigantea*) from the other three stony corals (*O. faveolata*, *S. pistillata*, and *A. digitifera*) happened 544 MYA (fig. 1A).

We also examined the differences of *Hox* (Homeobox) genes between the soft and stony corals. *Hox* genes encode transcription factors that perform diverse roles during development (Akam 1995). They are best known to define body plan (Akam 1995). We found the three stony corals have a similar and familiar pattern of *Hox* gene clusters (Ying et al. 2018) (fig. 1B). However, *Evx*, which is a member of the Antp superclass of *Hox* genes (Patel and Prince 2000), is absent in *D. gigantea* (fig. 1B) a finding that should be verified experimentally.

Here, we present a high-quality, draft genome of *Dendronephthya gigantea*, the first nonsymbiotic octocoral. Our analyses show the octocoral is the earliest diverged group among anthozoans with an estimated divergence time of 544 MYA from the hexacorals. It adds a new octocoral assembly for cnidarians, in addition to hexacoral and hydra genomes, thus it facilitates in depth comparative analyses of stony and soft corals that are either symbiotic and/or nonsymbiotic. The *D. gigantea* genome will support future experiments aimed at determining differences in the genetic

coping mechanisms between soft and stony corals in terms of calcification and survival strategies in the face of global warming and ocean acidification.

Materials and Methods

Genome Assembly and Annotation

A detailed description of the sample collection, DNA extraction, RNA extraction, genome size estimation, de novo genome assembly, and genome annotation can be found in the Supplementary Material online. In brief, PacBio long reads were used for a draft assembly processed by FALCON (version 0.3.0) (Chin et al. 2016) complemented by Illumina short paired-end reads for error-correction. We mapped Illumina short paired-end reads to the genome assembly to confirm the high quality using BWA (version 0.7.12) (Li, 2013) resulting in a 95.9% mapping rate. For gene prediction, we merged ab initio- and homology-based predictions using AUGUSTUS (version 3.1) (Stanke et al. 2008) with additional information obtained from homology-based predicted *D. gigantea* gene models, RNA-seq data of the planula and polyp of *D. gigantea* and polyps of *Scleronephthya gracillimum* (unpublished data), and Expressed Sequence Tags (ESTs) of corals downloaded from NCBI database.

Phylogenetic Analysis and *Hox* Gene Clusters Identification

We examined orthologous gene clustering of complete protein-coding genes from the six published cnidarians (*Orbicella faveolata*, *Stylophora pistillata*, *Acropora digitifera*, *Nematostella vectensis*, *Aiptasia pallida*, and *Hydra magnipapillata*) and seven noncnidarian metazoans (*Danio rerio*, *Homo Sapiens*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Trichoplax adhaerens*, *Amphimedon queenslandica*, and *Mnemiopsis leidyi*). Our outgroup was the unicellular holozoan, *Monosiga brevicollis*. Clusters were generated using OrthoMCL (version 2.0.9) (Li et al. 2003) with an E-value cutoff of 1E-20.

We estimated the phylogeny using 197 single copy orthologs using the PROTGAMMAJTT model in RAXML (version 8.2.8) (Stamatakis 2014). The divergence times were

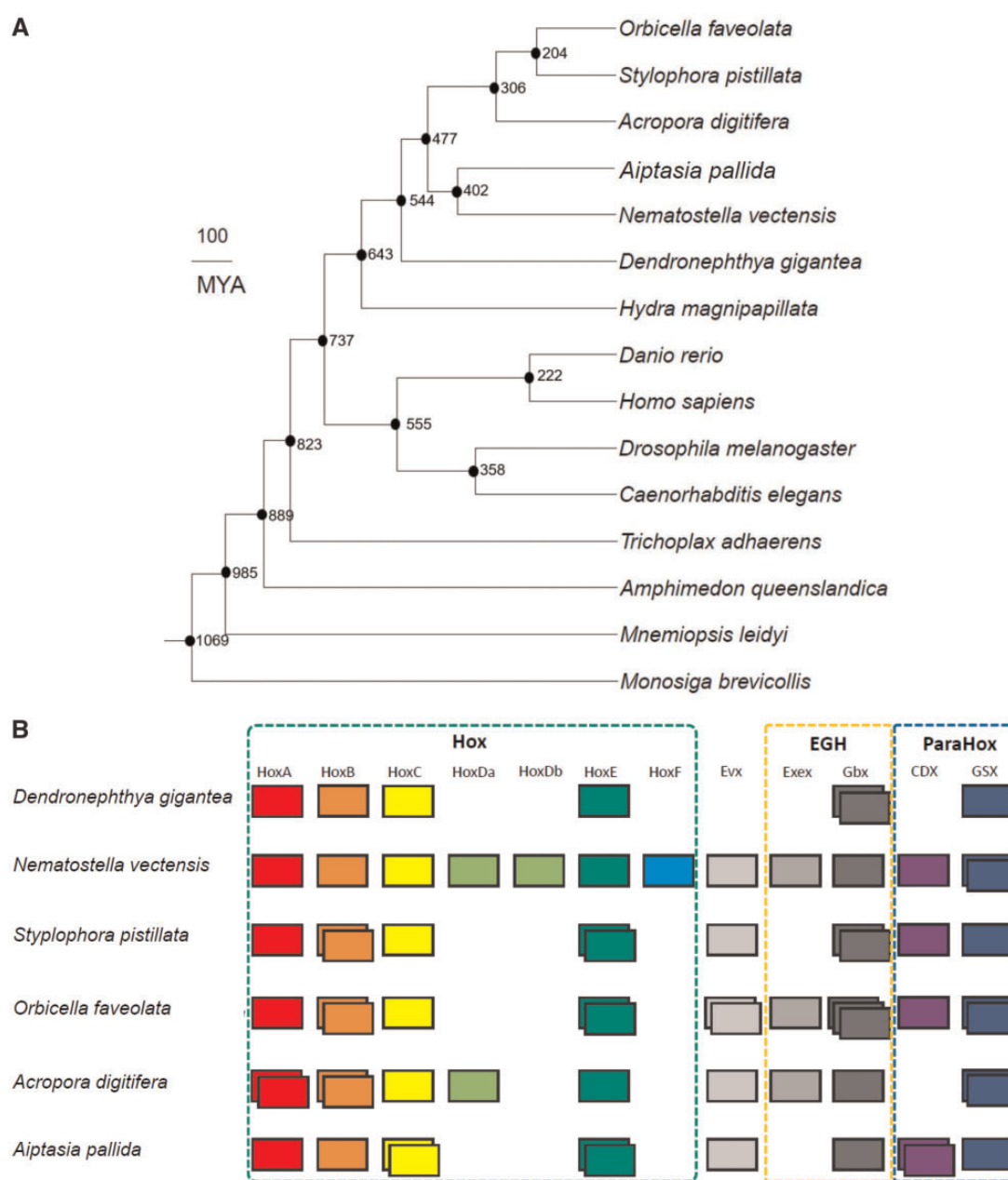


Fig. 1.—Phylogenetic relationship and *Hox* gene clusters of *Dendronephthya gigantea* and other species. (A) Tree shows the phylogeny with divergence time among 15 species. Numbers in each branch denote the estimated divergence time (MYA). (B) Green dashed-line box denotes *Hox* gene cluster (*HoxA*, *HoxB*, *HoxC*, *HoxDa*, *HoxDb*, *HoxE*, and *HoxF*), yellow dashed-line box denotes *EGF* gene cluster (*Evx* and *Gbx*), and blue dashed-line box denotes *ParaHox* gene cluster (*CDX* and *GSX*). The number of boxes shows the number of each gene copies in the genome.

estimated using the MCMCtree program in PAML package (version 4.8) (Yang 2007) with the independent rates model (clock = 2). The date of the node between *D. melanogaster*–*C. elegans* was constrained to 743 MYA and *H. sapiens*–*D. rerio* was constrained to 435 MYA based on the TimeTree database (Kumar et al. 2017).

To identify and classify *Hox* gene cluster patterns, we sought for all instances of the homeobox domain based on

Pfam database (Finn et al. 2016) using HMMER (version 3.1b2) (Finn et al. 2011) and InterProScan (version 5.32-71.0) (Jones et al. 2014; Mitchell et al. 2018). Homeobox domain genes were classified using BLAST (version 2.2.28) (Altschul et al. 1990) against HomeoDB (Zhong et al. 2008; Zhong and Holland 2011) and mapping to the homeobox domain of *N. vectensis* *Hox* genes from GenBank (Lipman, et al. 2016).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was supported by a grant from the Marine Biotechnology Program (20170305, Development of Biomedical materials based on marine proteins) and the Collaborative Genome Program (20180430) funded by the Ministry of Oceans and Fisheries, Korea and the Genome Korea Project in Ulsan Research Funds (1.180024.01 and 1.180017.01) of Ulsan National Institute of Science & Technology (UNIST). We also thank Genome Research Foundation for research support and the Korea Institute of Science and Technology Information (KISTI) for providing us access to the Korea Research Environment Open NETwork (KREONET).

Literature cited

- Akam M. 1995. *Hox* genes and the evolution of diverse body plans. *Philos Trans R Soc Lond B Biol Sci.* 349(1329):313–319.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215(3):403–410.
- Baumgarten S, et al. 2015. The genome of *Aiptasia*, a sea anemone model for coral symbiosis. *Proc Natl Acad Sci U S A.* 112(38):11893–11898.
- Carpenter KE, et al. 2008. One-Third of Reef-Building Corals Face Elevated Extinction Risk from Climate Change and Local Impacts. *science* 321: 560–563. doi: 10.1126/science.1159196.
- Chin C-S, et al. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods.* 13(12):1050.
- de Paula AF, Creed JC. 2004. Two species of the coral *Tubastraea* (Cnidaria, Scleractinia) in Brazil: a case of accidental introduction. *Bull Mar Sci.* 74:175–183.
- de Putron SJ, McCorkle DC, Cohen AL, Dillon A. 2011. The impact of seawater saturation state and bicarbonate ion concentration on calcification by new recruits of two Atlantic corals. *Coral Reefs* 30:321–328.
- Ellegren H, Galtier N. 2016. Determinants of genetic diversity. *Nat Rev Genet.* 17(7):422.
- Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39(Suppl):W29–W37.
- Finn RD, et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44(D1):D279–D285.
- Friedlander AM, Parrish JD. 1998. Habitat characteristics affecting fish assemblages on a Hawaiian coral reef. *J Exp Mar Biol Ecol.* 224:1–30.
- Gili J-M, Coma R. 1998. Benthic suspension feeders: their paramount role in littoral marine food webs. *Trends Ecol Evol.* 13(8):316–321.
- Hoegh-Guldberg O. 1999. Climate change, coral bleaching and the future of the world's coral reefs. *Mar Freshwater Res.* 50:839–866.
- Hwang S-J, Song J-I. 2007. Reproductive biology and larval development of the temperate soft coral *Dendronephthya gigantea* (Alcyonacea: Nephtheidae). *Mar Biol.* 152:273–284.
- Imbs AB, et al. 2007. Comparison of fatty acid compositions of azooxanthellate *Dendronephthya* and zooxanthellate soft coral species. *Comp Biochem Physiol B Biochem Mol Biol.* 148(3):314–321.
- Inoue S, Kayanne H, Yamamoto S, Kurihara H. 2013. Spatial community shift from hard to soft corals in acidified water. *Nat Climate Change.* 3:683.
- Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9):1236–1240.
- Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol.* 34(7):1812–1819.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv Preprint arXiv 1303.3997.*
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13(9):2178–2189.
- Lipman DJ, et al. 2016. GenBank. *Nucleic acids research* 45: D37–D42. doi: 10.1093/nar/gkw1070.
- Liu B, et al. 2013. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *arXiv Preprint arXiv 1308.2012.*
- Lopes AR, et al. 2018. Physiological resilience of a temperate soft coral to ocean warming and acidification. *Cell Stress and Chaperones* 23: 1093–1100. doi: 10.1007/s12192-018-0919-9.
- Mitchell AL, et al. 2018. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic acids research* 47: D351–D360. doi: 10.1093/nar/gky1100.
- Pandolfi JM, Connolly SR, Marshall DJ, Cohen ALJs. 2011. Projecting coral reef futures under global warming and ocean acidification. *Science* 333:418–422.
- Patel NH, Prince VE. 2000. Beyond the *Hox* complex. *Genome Biol.* 1(5):reviews1027–reviews1021.
- Putnam NH, et al. 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317(5834):86–94.
- Ries J, Cohen A, McCorkle D. 2010. A nonlinear calcification response to CO₂-induced ocean acidification by the coral *oculina* *Arbuscula*. *Coral Reefs* 29:661–674.
- Santodomingo N, et al. 2013. Diversity and distribution of azooxanthellate corals in the Colombian Caribbean. *Mar Biodivers.* 43:7–22.
- Shinzato C, et al. 2011. Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* 476(7360):320.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.
- Snelling J, Dziedzic K, Guemond S, Meyer E 2017a. Development of an integrated genomic map for a threatened Caribbean coral (*Orbicella faveolata*). *bioRxiv*: 183467. doi: 10.1101/183467.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24(5):637–644.
- van de Water JA, Allemand D, Ferrier-Pagès C. 2018. Host-microbe interactions in octocoral holobionts-recent advances and perspectives. *Microbiome* 6:64.
- Voolstra CR, et al. 2017. Comparative analysis of the genomes of *stylophora pistillata* and *acropora digitifera* provides evidence for extensive differences between species of corals. *Sci Rep.* 7(1):17583.
- Waterhouse RM, et al. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol.* 35(3):543–548.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.
- Ying H, et al. 2018. Comparative genomics reveals the distinct evolutionary trajectories of the robust and complex coral lineages. *Genome Biol.* 19:175.
- Zhong YF, Butts T, Holland PW. 2008. HomeoDB: a database of homeobox gene diversity. *Evol Dev.* 10(5):516–518.
- Zhong Y, Holland PW. 2011. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evol Dev.* 13(6):567–568.

Associate editor: Howard Ochman